ISSN 2616-5902 Vol. 4, Issue 5: 10-14, DOI: 10.25236/AJBM.2022.040503

# Forecast of Shanghai Stock Exchange 50 Stock Index Based on CNN-LSTM

**Qianxiao Fei, Yuwei Xiang**

*SHIEN-MING WU School of Intelligent Engineering, South China University of Technology, Guangzhou, Guangdong, 510000, China*

***Abstract:*** *Stock is one of the major financial management methods in today's society, and its index prediction method is widely on research. Stock index is regarded as an important stock market analysis indicators, precise prediction and analysis of stock index can effectively reflect the current stock fair situation. This paper summarizes the existing forecasting methods in the current academic research, the use of relevant data to analyze the accuracy of the newly established model and according to the prediction characteristics and prediction objects of different forecasting methods, the deep neural network is applied to the prediction of stock index. Taking the Shanghai Stock Exchange 50 (SSE 50) stock index as the research object, long short term memory (LSTM)and Convolutional Neural Networks-Long short term memory (CNN-LSTM) helps to established a high accuracy stock index prediction model. The static and rolling data prediction results show that CNN-LSTM model has high accuracy in prediction and is more suitable for investors to use in investment decisions.*

***Keyword:*** *CNN; LSTM; CNN-LSTM; Stock Index Prediction*

## 1. Introduction

The market index matters since it is a major symbol that reflect market conditions. The Shanghai Stock Exchange, as one of the most influential index in the Chinese share market, takes all the stocks listed on the Shanghai Stock Exchange (SSE). The stock forecast is helpful for us to make the strategy efficiently. How to forecast the Shanghai Stock Index through the prediction model has become an important topic for scholars today.

Based on the data characteristics of financial market, its prediction method commonly adopts time series model to get the predicted value and machine learning correlation algorithm. At present, the prediction methods of time series mainly use (autoregressive integrated moving average) ARIMA and (Generalized Autoregressive conditional heteroskedasticity) GARCH for the characteristics of market index. However, due to the typical influence of the characteristics of the financial market and the feature of high noise, uncertainty, nonlinearity and non-stationarity of the basic data, these characteristics make it difficult to accurately forecast the traditional time series prediction model.

At present, the prediction method of machine learning adopts artificial neural network (ANN), support vector machine (SVM), short and long-term memory (LSTM) and other methods, and raise the prediction efficiency of the SSE index data by combining the prediction of multiple value of market indicators and adding hidden units to reduce the error. In the paper [1], ANN is combined with nonlinear optimization technology to greatly improve its prediction accuracy, but it still has the characteristics of slow convergence. In the paper [2], LSTM model is proposed to deal the problem of gradient explosion which occurs in traditional neural network and recurrent neural network through the hidden element structure in the model features.

In order to obtain the prediction accuracy that network structure runs in the case of limited data, this paper intends to use the fusion of convolutional neural network CNN and LSTM, and design the fusion of CNN-LSTM short-term stock index prediction model. LSTM can be combined with hidden units to improve and replace the structure, and convolutional neural network CNN can effectively extract local features. Therefore, CNN-LSTM network is adopted as the prediction model and predict the market index, and a relatively high accuracy is obtained by comparing it with LSTM.

ISSN 2616-5902 Vol. 4, Issue 5: 10-14, DOI: 10.25236/AJBM.2022.040503

## 2. Forecast of SSE 50 stock index based on CNN-LSTM

### *2.1. Prediction model of convolutional neural network*

As one of the most commonly used neural network structures, convolutional neural network is mainly composed of input layer, convolutional layer, pooling layer, full connection layer and output layer. It has the characteristics of local convolution, weight sharing and pooled dimension reduction.

Local convolution is used to cover only a small part of the data of the convolution kernel, and multiple convolution kernels are used to obtain the high-dimensional local features of the matrix, so that local features of the data can be extracted well. The weight sharing feature makes the neurons in the hidden layer of the network share the same parameters, thus having fewer free parameters, thus reducing the number of parameters, avoiding the over-fitting problem, improving the generalization ability, and better improving the training speed of the neural network. The pooling layer of convolutional neural network has the function of dimensionality reduction, namely convolution dimensionality reduction, to extract high-level features on the basis that there is no loss in original data features.

Its powerful feature extraction and fast training speed can be widely used in time series data classification tasks. In this paper, CNN model's strong classification ability is used to predict stock price fluctuations. Since the traditional convolutional neural network has the requirement that the number of rows and columns of input data is equal, stock price data cannot meet this condition, so the input data are screened and processed in this paper.

In the input layer, "V1", "V2", "V3", "V4", "V5", "V6" and "V7" are respectively the opening price, the highest price, the lowest price, the closing price, the pre-closing price, the trading volume, the rise and fall. The output result is composed of multiple feature planes, which correspond to the convolution kernel one by one, and the value of each neuron in the feature plane is calculated by the corresponding convolution kernel.

### *2.2. LSTM*

LSTM is an improved cyclic neural network with four layers: input layer, LSTM layer, full connection layer and output layer. The model added forgetting neurons and memory neurons on the basis of RNN to increase the time complexity of the hidden layer. LSTM's three-gate structure, including input gate, forgetting gate and output gate, slows down the phenomenon of gradient disappearance. Its structure principle is shown in Figure 1:

In view of the nonlinear and stochastic characteristics of short-term stock forecast, this paper based on LSTM short-term stock forecast model, in order to fully explore the characteristics of stock data in the time dimension of the day, improve the accuracy of short-term customer flow forecast.

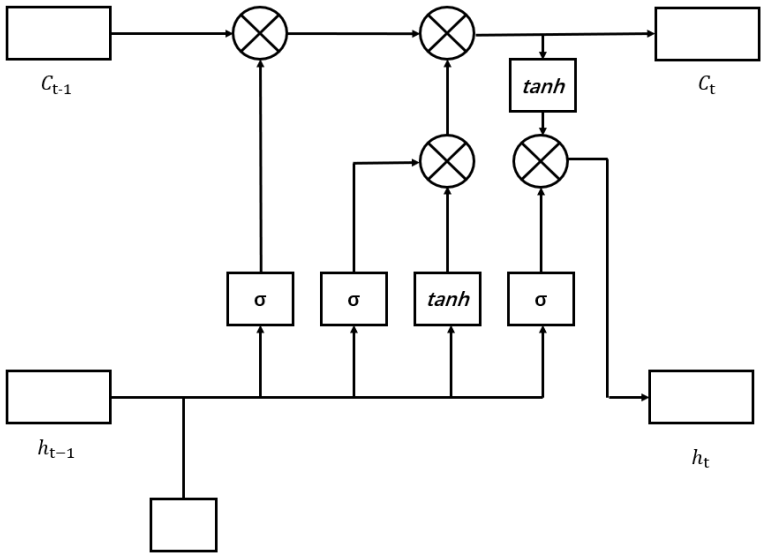



*Figure 1: The principle of LSTM*

ISSN 2616-5902 Vol. 4, Issue 5: 10-14, DOI: 10.25236/AJBM.2022.040503

### *2.3. CNN-LSTM*

The process of CNN-LSTM model for stock index prediction is divided into 2 parts. The first half is CNN network layer, which is used for spatial feature extraction of stock related data. CNN features are used to help realize feature extraction of stock features, and parameter sharing features are used to reduce the number of network parameters. Pooling layer is used for dimensionality reduction to achieve the purpose of greatly reducing the size of the convolution kernel while retaining the characteristics of the convolution kernel. Its structural model is shown in Figure 2:

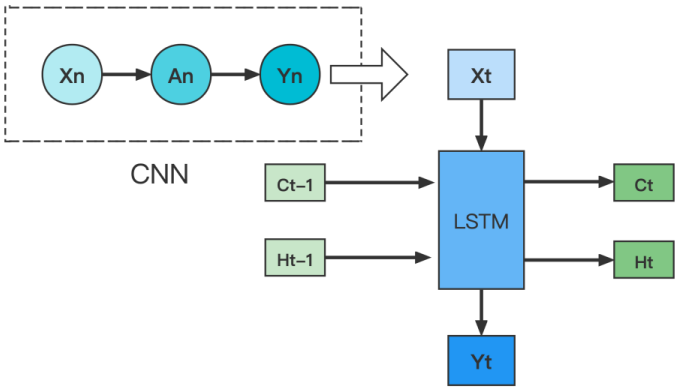



*Figure 2: The principle of LSTM*

## 3. Analysis of empirical results

### *3.1. Data sources and data preprocessing*

In order to evaluate the accuracy of CNN-LSTM neural network in stock index prediction, the DATA of SSE 5 stock index from February 26, 2015 to December 16, 2021 which has totally 1639 group of data is selected from choice database to predict the stock index. sample in days, and some missing values were deleted. After data analysis, it is planned to select the opening price, highest price, lowest price, closing price, advance price, trading volume and rise or fall as the input data to normalize the data, and the specific formula is as follows:

$$x'_j = \frac{\max(x_j) - x_j}{\max(x_j) - \min(x_j)} \tag{1}$$

$$x_j = The\ value\ of\ each\ field\ in\ the\ data$$

A total of 1639 sets of stock index data were obtained. The above data comes from the Choice database.

### *3.2. Model solving and model evaluation*

In the parameter data of machine learning in this paper, the training data accounts for 0.8 of the total data, that is, 1331 sets of data are used for training, the remaining 328 sets of data are used for prediction, the batch sample size batch size is 24, and the maximum number of iterations max epochs = 60.

The specific steps are as follows:

(1) Obtain historical data of SSE 50 stocks, including seven fields such as opening price, closing price and trading volume. Data normalization, elimination, missing values.

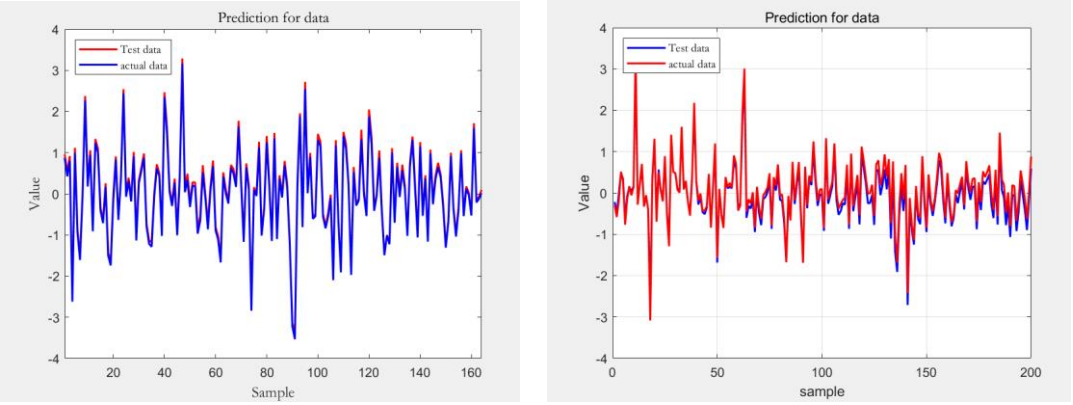
(2) The normalized data set is used to extract local features through CNN first. In order to prevent information loss, the operation of pooling layer is not carried out in the training, and the feature extraction results are directly input to LSTM, and then LSTM is used to extract the long-distance features of these local features, and then the transformation is input to the full connection layer. Based on the comparison between the training prediction results and the real value, the prediction accuracy of CNN-LSTM is given.

(3) The same predicted data are brought into the LSTM model for training, and the proportion of the training and predicted data remains unchanged. The prediction accuracy can be obtained by MAPE and

ISSN 2616-5902 Vol. 4, Issue 5: 10-14, DOI: 10.25236/AJBM.2022.040503

RMSE calculation between the obtained results and the real value.

(4) According to the results of the two neural network prediction models, the accuracy of the two methods for stock index prediction is compared, and the neural network prediction model more suitable for investors is selected. While keeping the input data unchanged, the data were respectively input into LSTM and CNN-LSTM models and retrograde training, and the training results are shown in Figure 3.




*a) CNN-LSTM* *b) LSTM*

*Figure 3: The result of LSTM and CNN-LSTM*

In order to investigate the effectiveness and performance of CNN-LSTM model, root mean square Error (RMSE) and mean absolute error percentage (MAPE) were selected to evaluate the prediction accuracy of CNN-LSTM model. The smaller RMSE and MAPE are, the higher the prediction accuracy of the model is. The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(X_{PRE,t} - X_{REAL,t})^2} \quad (2)$$

$$MAPE = (\frac{N}{100})\sum_{t=1}^{N}|\frac{X_{PRE,t}-X_{REAL,t}}{X_{REAL,t}}| \quad (3)$$

$X_{PRE,t} = The\ predicted\ value\ of\ the\ stock\ outcome$,

$X_{REAL,t} = The\ real\ value\ of\ the\ stock\ outcome$, N is the number of sample.

The mean absolute error percentage MAPE of CNN-LSTM was 0.5430, the mean absolute error percentage MAPE of LSTM was 0.6762, the root mean square error of CNN-LSTM was 0.2121, and the root mean square error of LSTM was 0.2897.

*Table 1: The error analysis of LSTM and CNN-LSTM*

| | LSTM | CNN-LSTM |
|---|---|---|
| RMSE | 0.2897 | 0.2121 |
| MAPE | 0.6762 | 0.5430 |

It can be seen from Table 1 that both LSTM and CNN-LSTM have certain prediction ability, but CNN-LSTM's prediction ability is better in terms of prediction accuracy and prediction error.

## 4. Conclusion

In this paper, seven indicators such as the stock opening price and closing price are introduced to predict the trend of the stock index, and the CNN-LSTM prediction model is constructed. The DATA of SSE 5 stock index from February 26, 2015 to December 16, 2021 are selected from choice database to predict the stock index. The static prediction results show that CNN-LSTM model has higher prediction accuracy than LSTM and other reference models, and has higher advantages in real-time prediction.

## References

*[1] BARBULESCU C, KILYENI S, DEACU A, et al. Artificial neural network based monthly load curves forecasting[C]//2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI). Timisoara, Romania. IEEE, 2016: 237–242.*

ISSN 2616-5902 Vol. 4, Issue 5: 10-14, DOI: 10.25236/AJBM.2022.040503

*[2] BIANCHI F M, MAIORINO E, KAMPFFMEYER M C, et al. Recurrent neural networks for short-term load forecasting [M]. Cham: Springer International Publishing, 2017.*
*[3] Di P L, Honchar O. Artificial Neural Networks Architectures for Stock Price Prediction: Comparisons and Applications [J]. International Journal of Circuits, Systems and Signal Processing, 2016, (10).*
*[4] Lee K B, Cheon S, Kim C O. A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes [J]. IEEE Transactions on Semiconductor Manufacturing, 2017, 30(2).*
*[5] Hochreiter S, Schmidhuber J. Long Short-term Memory [J]. Neural Computation, 1997, (8).*